moments are immediately seen to be zero by symmetry and the even moments are computed using the following theorem. In all cases it is assumed that the process begins indefinitely far in the past with $2r$ finite initial moments.

THEOREM 1: *For integer r, the $2r$th moment of a first-order linear ARCH process with $\alpha_0 > 0$, $\alpha_1 \geq 0$, exists if, and only if,*

$$\alpha_1^r \prod_{j=1}^{r} (2j-1) < 1.$$

A constructive expression for the moments is given in the proof.

PROOF: See Appendix.

The theorem is easily used to find the second and fourth moments of a first-order process. Letting $w_t = (y_t^4, y_t^2)'$,

$$E(w_t \mid \psi_{t-1}) = \begin{pmatrix} 3\alpha_0^2 \\ \alpha_0 \end{pmatrix} + \begin{pmatrix} 3\alpha_1^2 & 6\alpha_0\alpha_1 \\ 0 & \alpha_1 \end{pmatrix} w_{t-1}$$

The condition for the variance to be finite is simply that $\alpha_1 < 1$, while to have a finite fourth moment it is also required that $3\alpha_1^2 < 1$. If these conditions are met, the moments can be computed from (A4) as

$$(15) \qquad E(w_t) = \begin{bmatrix} \left[\dfrac{3\alpha_0^2}{(1-\alpha_1)^2} \right] \left[\dfrac{1-\alpha_1^2}{1-3\alpha_1^2} \right] \\ \dfrac{\alpha_0}{1-\alpha_1} \end{bmatrix}.$$

The lower element is the unconditional variance, while the upper product gives the fourth moment. The first expression in square brackets is three times the squared variance. For $\alpha_1 \neq 0$, the second term is strictly greater than one implying a fourth moment greater than that of a normal random variable.

The first-order ARCH process generates data with fatter tails than the normal density Many statistical procedures have been designed to be robust to large errors, but to the author's knowledge, none of this literature has made use of the fact that temporal clustering of outliers can be used to predict their occurrence and minimize their effects. This is exactly the approach taken by the ARCH model

4 GENERAL ARCH PROCESSES

The conditions for a first-order linear ARCH process to have a finite variance and, therefore, to be covariance stationary can directly be generalized for pth-order processes.

THEOREM 2: *The pth-order linear ARCH processes, with $\alpha_0 > 0$, $\alpha_1, \ldots, \alpha_p \geq 0$, is covariance stationary if, and only if, the associated characteristic equation has all roots outside the unit circle. The stationary variance is given by $E(y_t^2) = \alpha_0/(1 - \sum_{j=1}^{p} \alpha_j)$.*

PROOF: See Appendix.

Although the pth-order linear model is a convenient specification, it is likely that other formulations of the variance model may be more appropriate for particular applications. Two simple alternatives are the exponential and absolute value forms:

(16) $$h_t = \exp(\alpha_0 + \alpha_1 y_{t-1}^2),$$

(17) $$h_t = \alpha_0 + \alpha_1 |y_{t-1}|.$$

These provide an interesting contrast. The exponential form has the advantage that the variance is positive for all values of alpha, but it is not difficult to show that data generated from such a model have infinite variance for any value of $\alpha_1 \neq 0$. The implications of this deserve further study. The absolute value form requires both parameters to be positive, but can be shown to have finite variance for any parameter values.

In order to find estimation results which are more general than the linear model, general conditions on the variance model will be formulated and shown to be implied for the linear process.

Let ξ_t be a $p \times 1$ random vector drawn from the sample space Ξ, which has elements $\xi_t' = (\xi_{t-1}, \ldots, \xi_{t-p})$. For any ξ_t, let ξ_t^* be identical, except that the mth element has been multiplied by -1, where m lies between 1 and p.

DEFINITION: The ARCH process defined by (1) and (3) is *symmetric* if

(a) $h(\xi_t) = h(\xi_t^*)$ for any m and $\xi_t \epsilon \Xi$,

(b) $\partial h(\xi_t)/\partial \alpha_i = \partial h(\xi_t^*)/\partial \alpha_i$ for any m, i and $\xi_t \epsilon \Xi$,

(c) $\partial h(\xi_t)/\partial \xi_{t-m} = -\partial h(\xi_t^*)/\partial \xi_{t-m}$ for any m and $\xi_t \epsilon \Xi$.

All the functions described have been symmetric. This condition is the main distinction between mean and variance models.

Another characterization of general ARCH models is in terms of regularity conditions

DEFINITION: The ARCH model defined by (1) and (3) is *regular* if

(a) $\min h(\xi_t) \geq \delta$ for some $\delta > 0$ and $\xi_t \epsilon \Xi$,

(b) $E(|\partial h(\xi_t)/\partial \alpha_i||\partial h(\xi_t)/\partial \xi_{t-m}| \, |\psi_{t-m-1})$ exists for all i, m, t.

The first portion of the definition is very important and easy to check, as it requires the variance always to be positive. This eliminates, for example, the log-log autoregression The second portion is difficult to check in some cases, yet should generally be true if the process is stationary with bounded derivatives, since conditional expectations are finite if unconditional ones are. Condition (b) is a sufficient condition for the existence of some expectations of the Hessian used in Theorem 4. Presumably weaker conditions could be found

THEOREM 3. *The pth-order linear ARCH model satisfies the regularity conditions, if* $\alpha_0 > 0$ *and* $\alpha_1, \ldots, \alpha_p \geq 0$

PROOF. See Appendix.

In the estimation portion of the paper, a very substantial simplification results if the ARCH process is symmetric and regular.

5 ARCH REGRESSION MODELS

If the ARCH random variables discussed thus far have a non-zero mean, which can be expressed as a linear combination of exogenous and lagged dependent variables, then a regression framework is appropriate, and the model can be written as in (4) or (5). An alternative interpretation for the model is that the disturbances in a linear regression follow an ARCH process.

In the pth-order linear case, the specification and likelihood are given by

$$
\begin{aligned}
& y_t \mid \psi_{t-1} \sim N(x_t\beta, h_t), \\
& h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_p\epsilon_{t-p}^2, \\
(18) \qquad & \epsilon_t = y_t - x_t\beta, \\
& l = \frac{1}{T}\sum_{t=1}^{T} l_t, \\
& l_t = -\tfrac{1}{2}\log h_t - \tfrac{1}{2}\epsilon_t^2/h_t,
\end{aligned}
$$

where x_t may include lagged dependent and exogenous variables and an irrelevant constant has been omitted from the likelihood. This likelihood function can be maximized with respect to the unknown parameters α and β. Attractive methods for computing such an estimate and its properties are discussed below

Under the assumptions in (18), the ordinary least squares estimator of β is still consistent as x and ϵ are uncorrelated through the definition of the regression as a conditional expectation. If the x's can be treated as fixed constants then the least squares standard errors will be correct; however, if there are lagged dependent variables in x_t, the standard errors as conventionally computed will not be consistent, since the squares of the disturbances will be correlated with squares of the x's. This is an extension of White's [18] argument on heteroscedasticity and it suggests that using his alternative form for the covariance matrix would give a consistent estimate of the least-squares standard errors.

If the regressors include no lagged dependent variables and the process is stationary, then letting y and x be the $T \times 1$ and $T \times K$ vector and matrix of dependent and independent variables, respectively,

$$E(y \mid x) = x\beta, \qquad \mathrm{Var}(y \mid x) = \sigma^2 I, \tag{19}$$

and the Gauss–Markov assumptions are statisfied. Ordinary least squares is the best linear unbiased estimator for the model in (18) and the variance estimates are unbiased and consistent. However, maximum likelihood is different and consequently asymptotically superior; ordinary least squares does not achieve the Cramer–Rao bound. The maximum-likelihood estimator is nonlinear and is more efficient than OLS by an amount calculated in Section 6.

The maximum likelihood estimator is found by solving the first order conditions. The derivative with respect to β is

$$\frac{\partial l_t}{\partial \beta} = \frac{\epsilon_t x_t'}{h_t} + \frac{1}{2h_t}\frac{\partial h_t}{\partial \beta}\left(\frac{\epsilon_t^2}{h_t} - 1\right). \tag{20}$$

The first term is the familiar first-order condition for an exogenous heteroscedastic correction; the second term results because h_t is also a function of the β's, as in Amemiya [1]. Substituting the linear variance function gives

$$\frac{\partial l}{\partial \beta} = \frac{1}{T}\sum\left[\frac{\epsilon_t x_t'}{h_t} - \frac{1}{h_t}\left(\frac{\epsilon_t^2}{h_t} - 1\right)\sum_j \alpha_j \epsilon_{t-j} x_{t-j}'\right], \tag{21}$$

which can be rewritten approximately by collecting terms in x and ϵ as

$$\begin{aligned}\frac{\partial l}{\partial \beta} &= \frac{1}{T}\sum_t x_t'\epsilon_t\left[h_t^{-1} - \sum_{j=1}^{p}\alpha_j h_{t+j}^{-2}\left(\epsilon_{t+j}^2 - h_{t+j}\right)\right] \\ &\approx \frac{1}{T}\sum_t x_t'\epsilon_t s_t.\end{aligned} \tag{22}$$

The Hessian is

$$\begin{aligned}\frac{\partial^2 l_t}{\partial \beta \partial \beta'} = &-\frac{x_t' x_t}{h_t} - \frac{1}{2h_t^2}\frac{\partial h_t}{\partial \beta}\frac{\partial h_t}{\partial \beta'}\left(\frac{\epsilon_t^2}{h_t}\right) \\ &- \frac{2\epsilon_t x_t'}{h_t^2}\frac{\partial h_t}{\partial \beta} + \left(\frac{\epsilon_t^2}{h_t} - 1\right)\frac{\partial}{\partial \beta'}\left[\frac{1}{2h_t}\frac{\partial h_t}{\partial \beta}\right].\end{aligned}$$

Taking conditional expectations of the Hessian, the last two terms vanish because h_t is entirely a function of the past. Similarly, ϵ_t^2/h_t becomes one, since it is the only current value in the second term. Notice that these results hold regardless of whether x_t includes lagged-dependent variables. The information matrix is the average over all t of the expected value of the conditional expectation and is, therefore, given by

$$\mathcal{I}_{\beta\beta} = \frac{1}{T}\sum_t E\left[E\left(\frac{\partial^2 l_t}{\partial\beta\partial\beta'} \mid \psi_{t-1}\right)\right] \tag{23}$$

$$= \frac{1}{T}\sum_t E\left[\frac{x_t'x_t}{h_t} + \frac{1}{2h_t^2}\frac{\partial h_t}{\partial\beta}\frac{\partial h_t}{\partial\beta'}\right].$$

For the pth order linear ARCH regression this is consistently estimated by

$$\hat{\mathcal{I}}_{\beta\beta} = \frac{1}{T}\sum\left[\frac{x_t'x_t}{h_t} + 2\sum_j \alpha_j^2 \frac{\epsilon_{t-j}^2}{h_t^2} x_{t-j}'x_{t-j}\right]. \tag{24}$$

By gathering terms in $x_t'x_t$, (24) can be rewritten, except for end effects, as

$$\hat{\mathcal{I}}_{\beta\beta} = \frac{1}{T}\sum_t x_t'x_t\left[h_t^{-1} + 2\epsilon_t^2\sum_{j=1}^{p}\alpha_j^2 h_{t+j}^{-2}\right] \tag{25}$$

$$\cong \frac{1}{T}\sum_t x_t'x_t r_t^2.$$

In a similar fashion, the off-diagonal blocks of the information matrix can be expressed as:

$$\mathcal{I}_{\alpha\beta} = \frac{1}{T}\sum_t E\left(\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\alpha}\frac{\partial h_t}{\partial\beta'}\right). \tag{26}$$

The important result to be shown in Theorem 4 below is that this off-diagonal block is zero. The implications are far-reaching in that estimation of α and β can be undertaken separately without asymptotic loss of efficiency and their variances can be calculated separately.

THEOREM 4: *If an ARCH regression model is symmetric and regular, then* $\mathcal{I}_{\alpha\beta} = 0$.

PROOF. See Appendix.

6 ESTIMATION OF THE ARCH REGRESSION MODEL

Because of the block diagonality of the information matrix, the estimation of α and β can be considered separately without loss of asymptotic efficiency.

Furthermore, either can be estimated with full efficiency based only on a consistent estimate of the other. See, for example, Cox and Hinkley [6, p. 308]. The procedure recommended here is to initially estimate β by ordinary least squares, and obtain the residuals. From these residuals, an efficient estimate of α can be constructed, and based upon these $\hat{\alpha}$ estimates, efficient estimates of β are found. The iterations are calculated using the scoring algorithm. Each step for a parameter vector ϕ produces estimates ϕ^{i+1} based on ϕ^i according to

$$(27) \qquad \phi^{i+1} = \phi^i + \left[\hat{\mathcal{I}}^i_{\phi\phi}\right]^{-1} \frac{1}{T} \sum_t \frac{\partial l_t^i}{\partial \phi},$$

where $\hat{\mathcal{I}}^i$ and $\partial l_t^i / \partial \phi$ are evaluated at ϕ^i. The advantage of this algorithm is partly that it requires only first derivatives of the likelihood function in this case and partly that it uses the statistical properties of the problem to tailor the algorithm to this application

For the pth-order linear model, the scoring step for α can be rewritten by substituting (12), (13), and (14) into (27) and interpreting y_t^2 as the residuals e_t^2. The iteration is simply

$$(28) \qquad \alpha^{i+1} = \alpha^i + (\tilde{z}'\tilde{z})^{-1}\tilde{z}'f^i$$

where

$$\tilde{z}_t = \left(1, e_{t-1}^2, \ldots, e_{t-p}^2\right)/h_t^i,$$

$$\tilde{z}' = (\tilde{z}'_1, \ldots, \tilde{z}'_T),$$

$$f_t^i = (e_t^2 - h_t^i)/h_t^i,$$

$$f^{i\prime} = (f_1^i, \ldots, f_T^i).$$

In these expressions, e_t is the residual from iteration i, h_t^i is the estimated conditional variance, and α^i is the estimate of the vector of unknown parameters from iteration i. Each step is, therefore, easily constructed from a least-squares regression on transformed variables The variance-covariance matrix of the parameters is consistently estimated by the inverse of the estimate of the information matrix divided by T, which is simply $2(\tilde{z}'\tilde{z})^{-1}$. This differs slightly from $\hat{\sigma}^2(\tilde{z}'\tilde{z})^{-1}$ computed by the auxiliary regression. Asymptotically, $\hat{\sigma}^2 = 2$, if the distributional assumptions are correct, but it is not clear which formulation is better in practice.

The parameters in α must satisfy some nonnegativity conditions and some stationarity conditions. These could be imposed via penalty functions or the parameters could be estimated and checked for conformity. The latter approach is used here, although a perhaps useful reformulation of the model might employ squares to impose the nonnegativity constraints directly:

$$(29) \qquad h_t = \alpha_0^2 + \alpha_1^2\epsilon_{t-1}^2 + \cdots + \alpha_p^2\epsilon_{t-p}^2.$$

Convergence for such an iteration can be formulated in many ways. Following Belsley [3], a simple criterion is the gradient around the inverse Hessian. For a parameter vector, ϕ, this is

$$\theta = \frac{\partial l'}{\partial \phi}\left(\frac{\partial^2 l}{\partial \phi \partial \phi'}\right)^{-1}\frac{\partial l}{\partial \phi}. \tag{30}$$

Using θ as the convergence criterion is attractive, as it provides a natural normalization and as it is interpretable as the remainder term in a Taylor-series expansion about the estimated maximum. In any case, substituting the gradient and estimated information matrix in (30), $\theta = R^2$ of the auxiliary regression.

For a given estimate of α, a scoring step can be computed to improve the estimate of beta. The scoring algorithm for β is

$$\beta^{i+1} = \beta^i + \left[\hat{\mathscr{I}}_{\beta\beta}\right]^{-1}\frac{\partial l^i}{\partial \beta}. \tag{31}$$

Defining $\tilde{x}_t = x_t r_t$ and $\tilde{e}_t = e_t s_t / r_t$ with $\tilde{x}$ and $\tilde{e}$ as the corresponding matrix and vector, (31) can be rewritten using (22) and (24) and e_t for the estimate of ϵ_t on the ith iteration, as

$$\beta^{i+1} = \beta^i + (\tilde{x}'\tilde{x})^{-1}\tilde{x}'\tilde{e} \tag{32}$$

Thus, an ordinary least-squares program can again perform the scoring iteration, and $(\tilde{x}'\tilde{x})^{-1}$ from this calculation will be the final variance-covariance matrix of the maximum likelihood estimates of β

Under the conditions of Crowder's [7] theorem for martingales, it can be established that the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ are asymptotically normally distributed with limiting distribution

$$\begin{aligned} &\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{D} N\left(0, \mathscr{I}_{\alpha\alpha}^{-1}\right), \\ &\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{D} N\left(0, \mathscr{I}_{\beta\beta}^{-1}\right). \end{aligned} \tag{33}$$

7 GAINS IN EFFICIENCY FROM MAXIMUM LIKELIHOOD ESTIMATION

The gain in efficiency from using the maximum-likelihood estimation rather than OLS has been asserted above. In this section, the gains are calculated for a special case. Consider the linear stationary ARCH model with $p = 1$ and all x_t exogenous. This is the case where the Gauss–Markov theorem applies and OLS has a variance matrix $\sigma^2(x'x)^{-1} = E\epsilon_t^2(\sum_t x_t' x_t)^{-1}$. The stationary variance is $\sigma^2 = \alpha_0/(1 - \alpha_1)$.

The information matrix for this case becomes, from (25),

$$E\left[\sum_t x_t' x_t\left(h_t^{-1} + 2\epsilon_t^2 \alpha_1^2 / h_{t+1}^2\right)\right].$$

With x exogenous, the expectation is only necessary over the scale factor. Because the disturbance process is stationary, the variance-covariance matrix is proportional to that for OLS and the relative efficiency depends only upon the scale factors. The relative efficiency of MLE to OLS is, therefore,

$$R = E\left(h_t^{-1} + 2\epsilon_t^2\alpha_1^2/h_{t+1}^2\right)\sigma^2.$$

Now substitute $h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2$, $\sigma^2 = \alpha_0/1 - \alpha_1$, and $\gamma = \alpha_1/1 - \alpha_1$. Recognizing that ϵ_{t-1}^2 and ϵ_t^2 have the same density, define for each

$$u = \epsilon\sqrt{(1-\alpha_1)/\alpha_0}\,.$$

The expression for the relative efficiency becomes

$$R = E\left(\frac{1+\gamma}{1+\gamma u^2}\right) + 2\gamma^2 E\frac{u^2}{\left(1+\gamma u^2\right)^2}\,, \tag{34}$$

where u has variance one and mean zero. From Jensen's inequality, the expected value of a reciprocal exceeds the reciprocal of the expected value and, therefore, the first term is greater than unity The second is positive, so there is a gain in efficiency whenever $\gamma \neq 0$. Eu^{-2} is infinite because u^2 is conditionally chi squared with one degree of freedom. Thus, the limit of the relative efficiency goes to infinity with γ:

$$\lim_{\gamma\to\infty} R \to \infty.$$

For α_1 close to unity, the gain in efficiency from using a maximum likelihood estimator may be very large.

8 TESTING FOR ARCH DISTURBANCES

In the linear regression model, with or without lagged-dependent variables, OLS is the appropriate procedure if the disturbances are not conditionally heteroscedastic. Because the ARCH model requires iterative procedures, it may be desirable to test whether it is appropriate before going to the effort to estimate it The Lagrange multiplier test procedure is ideal for this as in many similar cases See, for example, Breusch and Pagan [**4**, **5**], Godfrey [**12**], and Engle [**9**].

Under the null hypothesis, $\alpha_1 = \alpha_2 \cdots = \alpha_p = 0$ The test is based upon the score under the null and the information matrix under the null Consider the ARCH model with $h_t = h(z_t\alpha)$, where h is some differentiable function which, therefore, includes both the linear and exponential cases as well as lots of others and $z_t = (1, e_{t-1}^2, \ldots, e_{t-p}^2)$ where e_t are the ordinary least squares residuals Under the null, h_t is a constant denoted h^0. Writing $\partial h_t/\partial\alpha = h'z_t'$, where h' is the scalar derivative of h, the score and information can be written as

$$\frac{\partial l}{\partial \alpha}\bigg|_0 = \frac{h'}{2h^0}\sum_t z_t'\left(\frac{e_t^2}{h^0} - 1\right) = \frac{h^{0\prime}}{2h^0} z' f^0,$$

$$\mathcal{I}_{\alpha\alpha}^0 = \frac{1}{2}\left(\frac{h^{0\prime}}{h^0}\right)^2 E z' z,$$

and, therefore, the LM test statistic can be consistently estimated by

$$\xi^* = \tfrac{1}{2} f^{0\prime} z (z'z)^{-1} z' f^0 \tag{35}$$

where $z' = (z_1', \ldots, z_T')$, f^0 is the column vector of

$$\left(\frac{e_t^2}{h^0} - 1\right).$$

This is the form used by Breusch and Pagan [4] and Godfrey [12] for testing for heteroscedasticity. As they point out, all reference to the h function has disappeared and, thus, the test is the same for any h which is a function only of $z_t\alpha$.

In this problem, the expectation required in the information matrix could be evaluated quite simply under the null; this could have superior finite sample performance A second simplification, which is appropriate for this model as well as the heteroscedasticity model, is to note that plim $f^{0\prime}f^0/T = 2$ because normality has already been assumed. Thus, an asymptotically equivalent statistic would be

$$\xi = T f^{0\prime} z (z'z)^{-1} z' f^0 / f^{0\prime} f^0 = TR^2 \tag{36}$$

where R^2 is the squared multiple correlation between f^0 and z. Since adding a constant and multiplying by a scalar will not change the R^2 of a regression, this is also the R^2 of the regression of e_t^2 on an intercept and p lagged values of e_t^2. The statistic will be asymptotically distributed as chi square with p degrees of freedom when the null hypothesis is true.

The test procedure is to run the OLS regression and save the residuals. Regress the squared residuals on a constant and p lags and test TR^2 as a χ_p^2 This will be an asymptotically locally most powerful test, a characterization it shares with likelihood ratio and Wald tests. The same test has been proposed by Granger and Anderson [13] to test for higher moments in bilinear time series.

9 ESTIMATION OF THE VARIANCE OF INFLATION

Economic theory frequently suggests that economic agents respond not only to the mean, but also to higher moments of economic random variables. In financial theory, the variance as well as the mean of the rate of return are determinants of portfolio decisions. In macroeconomics, Lucas [16], for example, argues that the variance of inflation is a determinant of the response to various shocks. Furthermore, the variance of inflation may be of independent interest as it is the unanticipated component which is responsible for the bulk of the welfare loss due to inflation. Friedman [11] also argues that, as high inflation will generally be associated with high variability of inflation, the statistical relationship between inflation and unemployment should have a positive slope, not a negative one as in the traditional Phillips curve.

Measuring the variance of inflation over time has presented problems to various researchers. Khan [14] has used the absolute value of the first difference of inflation while Klein [15] has used a moving variance around a moving mean. Each of these approaches makes very simple assumptions about the mean of the distribution, which are inconsistent with conventional econometric approaches. The ARCH method allows a conventional regression specification for the mean function, with a variance which is permitted to change stochastically over the sample period. For a comparison of several measures for U.S. data, see Engle [10].

A conventional price equation was estimated using British data from 1958-II through 1977-II. It was assumed that price inflation followed wage increases; thus the model is a restricted transfer function

Letting $\dot{p}$ be the first difference of the log of the quarterly consumer price index and w be the log of the quarterly index of manual wage rates, the model chosen after some experimentation was

$$\dot{p} = \beta_1\dot{p}_{-1} + \beta_2\dot{p}_{-4} + \beta_3\dot{p}_{-5} + \beta_4(p - w)_{-1} + \beta_5. \tag{37}$$

The model has typical seasonal behavior with the first, fourth, and fifth lags of the first difference. The lagged value of the real wage is the error correction mechanism of Davidson, et al. [8], which restricts the lag weights to give a constant real wage in the long run. As this is a reduced form, the current wage rate cannot enter.

The least squares estimates of this model are given in Table I. The fit is quite good, with less than 1 per cent standard error of forecast, and all t statistics greater than 3. Notice that $\dot{p}_{-4}$ and $\dot{p}_{-5}$ have equal and opposite signs, suggesting that it is the acceleration of inflation one year ago which explains much of the short-run behavior in prices.

TABLE I
ORDINARY LEAST SQUARES (36)[a]

Variable	$\dot{p}_{-1}$	$\dot{p}_{-4}$	$\dot{p}_{-5}$	$(p-w)_{-1}$	Const	α_0 ($\times 10^{-6}$)	α_1
Coeff.	0 334	0 408	− 0.404	− 0 0559	0 0257	89	0
St. Err.	0.103	0 110	0.114	0.0136	0 00572		
t Stat	3 25	3 72	3.55	4 12	4 49		

[a] Dependent variable $\dot{p} = \log(P) - \log(P_{-1})$ where P is quarterly U K consumer price index $w = \log(W)$ where W is the U K. index of manual wage rates Sample period 1958-II to 1977-II

To establish the reliability of the model by conventional criteria, it was tested for serial correlation and for coefficient restrictions. Godfrey's [12] Lagrange multiplier test, for serial correlation up to sixth order, yields a chi-squared statistic with 6 degrees of freedom of 4 53, which is not significant, and the square of Durbin's h is 0.57. Only the 9th autocorrelation of the least squares residuals exceeds two asymptotic standard errors and, thus, the hypothesis of white noise disturbances can be accepted. The model was compared with an unrestricted regression, including all lagged p and w from one quarter through six. The asymptotic F statistic was 2.04, which is not significant at the 5 per cent level. When (37) was tested for the exclusion of w_{-1} through w_{-6}, the statistic was 2.34, which is barely significant at the 5 per cent but not the 2.5 per cent level. The only variable which enters significantly in either of these regressions is w_{-6} and it seems unattractive to include this alone.

The Lagrange multiplier test for a first-order linear ARCH effect for the model in (37) was not significant. However, testing for a fourth-order linear ARCH process, the chi-squared statistic with 4 degrees of freedom was 15 2, which is highly significant. Assuming that agents discount past residuals, a linearly declining set of weights was formulated to give the model

$$h_t = \alpha_0 + \alpha_1(0.4\epsilon_{t-1}^2 + 0.3\epsilon_{t-2}^2 + 0.2\epsilon_{t-3}^2 + 0.1\epsilon_{t-4}^2) \tag{38}$$

which is used in the balance of the paper. A two-parameter variance function was chosen because it was suspected that the nonnegativity and stationarity constraints on the α's would be hard to satisfy in an unrestricted model. The chi-squared test for $\alpha_1 = 0$ in (38) was 6.1, which has one degree of freedom.

One step of the scoring algorithm was employed to estimate model (37) and (38). The scoring step on α was performed first and then, using the new efficient $\hat{\alpha}$, the algorithm obtains in one step, efficient estimates of β. These are given in Table II. The procedure was also iterated to convergence by doing three steps on α, followed by three steps on β, followed by three more steps on α, and so forth. Convergence, within 0.1 per cent of the final value, occurred after two sets of α and β steps. These results are given in Table III.

The maximum likelihood estimates differ from the least squares effects primarily in decreasing the sizes of the short-run dynamic coefficients and increasing

TABLE II
MAXIMUM LIKELIHOOD ESTIMATES OF ARCH MODEL (36) (37)
ONE-STEP SCORING ESTIMATES[a]

Variable	p_{-1}	p_{-4}	p_{-5}	$(p-w)_{-1}$	Const.	$\alpha_0\ (\times 10^{-6})$	α_1
Coeff	0 210	0.270	– 0 334	– 0 0697	0 0321	19	0.846
St Err.	0.110	0.094	0 109	0.0117	0 00498	14	0.243
t Stat.	1 90	2.86	3 06	5 98	6 44	1.32	3 49

[a] Dependent variable $p = \log(P) - \log(P_{-1})$ where P is quarterly U K. consumer price index $w = \log(W)$ where W is the U K index of manual wage rates Sample period 1958-II to 1977-II.

TABLE III

MAXIMUM LIKELIHOOD ESTIMATES OF ARCH MODEL (36) (37)
ITERATED ESTIMATES[a]

Variables	p_{-1}	p_{-4}	$\dot{p}_{-5}$	$(p-w)_{-1}$	Const	$\alpha_0\ (\times 10^{-6})$	α_1
Coeff	0.162	0 264	−0.325	−0 0707	0.0328	14	0 955
St. Err	0 108	0 0892	0 0987	0.0115	0 00491	8 5	0 298
t Stat	1 50	2.96	3.29	6 17	6 67	1 56	3.20

[a] Dependent variable $p = \log(P) - \log(P_{-1})$ where P is quarterly U K consumer price index $w = \log(W)$ where W is the U K index of manual wage rates Sample period 1958-II to 1977-II

the coefficient on the long run, as incorporated in the error correction mechanism. The acceleration term is not so clearly implied as in the least squares estimates. These seem reasonable results, since much of the inflationary dynamics are estimated by a period of very severe inflation in the middle seventies. This, however, is also the period of the largest forecast errors and, hence, the maximum likelihood estimator will discount these observations. By the end of the sample period, inflationary levels were rather modest and one might expect that the maximum likelihood estimates would provide a better forecasting equation.

The standard errors for ordinary least squares are generally greater than for maximum likelihood. The least squares standard errors are 15 per cent to 25 per cent greater, with one exception where the standard error actually falls by 5 per cent to 7 per cent. As mentioned earlier, however, the least squares estimates are biased when there are lagged dependent variables. The Wald test for $\alpha_1 = 0$ is also significant.

The final estimates of h_t are the one-step-ahead forecast variances. For the one-step scoring estimator, these vary from 23×10^{-6} to 481×10^{-6}. That is, the forecast standard deviation ranges from 0 5 per cent to 2 2 per cent, which is more than a factor of 4. The average of the h_t, since 1974, is 230×10^{-6}, as compared with 42×10^{-6} during the last four years of the sixties. Thus, the standard deviation of inflation increased from 0.6 per cent to 1.5 per cent over a few years, as the economy moved from the rather predictable sixties into the chaotic seventies.

In order to determine whether the confidence intervals arising from the ARCH model were superior to the least squares model, the outliers were examined. The expected number of residuals exceeding two (conditional) standard deviations is 3 5. For ordinary least squares, there were 5 while ARCH produced 3. For least squares these occurred in '74-I, '75-I, '75-II, '75-IV, and '76-II; they all occur within three years of each other and, in fact, three of them are in the same year. For the ARCH model, they are much more spread out and only one of the least squares points remains an outlier, although the others are still large. Examining the observations exceeding one standard deviation shows similar effects. In the seventies, there were 13 OLS and 12 ARCH residuals outside one sigma, which are both above the expected value of 9. In the sixties, there were 6 for OLS, 10 for ARCH and an expected number of 12. Thus, the number of outliers for ordinary least squares is reasonable; however, the timing of their occurrence is far from random. The ARCH model comes closer to truly random residuals after standardizing for their conditional distributions.

This example illustrates the usefulness of the ARCH model for improving the performance of a least squares model and for obtaining more realistic forecast variances

University of California, San Diego

Manuscript received July, 1979; final revision received July, 1981

APPENDIX

PROOF OF THEOREM 1 Let

(A2) $$w_t' = (y_t^{2r}, y_t^{2(r-1)}, \ldots, y_t^2).$$

First, it is shown that there is an upper triangular $r \times r$ matrix A and $r \times 1$ vector b such that

(A2) $$E(w_t \mid \psi_{t-1}) = b + Aw_{t-1}$$

For any zero-mean normal random variable u, with variance σ^2,

$$E(u^{2r}) = \sigma^{2r} \prod_{j=1}^{r} (2j-1)$$

Because the conditional distribution of y is normal

(A3) $$E(y_t^{2m} \mid \psi_{t-1}) = h_t^{2m} \prod_{j=1}^{m} (2j-1) = (\alpha_1 y_{t-1}^2 + \alpha_0)^m \prod_{j=1}^{m} (2j-1).$$

Expanding this expression establishes that the moment is a linear combination of w_{t-1} Furthermore, only powers of y less than or equal to $2m$ are required, therefore, A in (A2) is upper triangular

Now

$$E(w_t \mid \psi_{t-2}) = b + A(b + Aw_{t-2})$$

or in general

$$E(w_t \mid \psi_{t-k}) = (I + A + A^2 + \cdots + A^{k-1})b + A^k w_{t-k}$$

Because the series starts indefinitely far in the past with $2r$ finite moments, the limit as k goes to infinity exists if, and only if, all the eigenvalues of A lie within the unit circle

The limit can be written as

$$\lim_{k\to\infty} E(w_t \mid \psi_{t-k}) = (I - A)^{-1}b,$$

which does not depend upon the conditioning variables and does not depend upon t. Hence, this is an expression for the stationary moments of the unconditional distribution of y.

(A4) $$E(w_t) = (I - A)^{-1}b$$

It remains only to establish that the condition in the theorem is necessary and sufficient to have all eigenvalues lie within the unit circle As the matrix has already been shown to be upper triangular, the diagonal elements are the eigenvalues From (A3), it is seen that the diagonal elements are simply

$$\alpha_1^m \prod_{j=1}^{m} (2j-1) = \prod_{j=1}^{m} \alpha_1(2j-1) \equiv \theta_m$$

for $m = 1, \ldots, r$. If θ_r exceeds or equals unity, the eigenvalues do not lie in the unit circle It must also be shown that if $\theta_r < 1$, then $\theta_m < 1$ for all $m < r$ Notice that θ_m is a product of m factors which are monotonically increasing If the mth factor is greater than one, then θ_{m-1} will necessarily be smaller than θ_m. If the mth factor is less than one, all the other factors must also be less than one and, therefore, θ_{m-1} must also have all factors less than one and have a value less than one. This establishes that a necessary and sufficient condition for all diagonal elements to be less than one is that $\theta_r < 1$, which is the statement in the theorem. *Q E D.*

PROOF OF THEOREM 2: Let

$$w_t' = (y_t^2, y_{t-1}^2, \ldots, y_{t-p}^2)$$

Then in terms of the companion matrix A,

(A5) $$E(w_t \mid \psi_{t-1}) = b + Aw_{t-1}$$

where $b' = (\alpha_0, 0, \ldots, 0)$ and

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdot & \alpha_p & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdot\cdot & 1 & 0 \end{bmatrix}.$$

Taking successive expectations

$$E(w_t \mid \psi_{t-k}) = (I + A + A^2 + \cdots + A^{k-1})b + A^k w_{t-k}.$$

Because the series starts indefinitely far in the past with finite variance, if, and only if, all eigenvalues lie within the unit circle, the limit exists and is given by

(A6) $$\lim_{k\to\infty} E(w_t \mid \psi_{t-k}) = (I - A)^{-1}b$$

As this does not depend upon initial conditions or on t, this vector is the common variance for all t As is well known in time series analysis, this condition is equivalent to the condition that all the roots of the characteristic equation, formed from the α's, lie outside the unit circle. See Anderson [2, p. 177] Finally, the limit of the first element can be rewritten as

(A7) $$Ey_t^2 = \alpha_0 \Big/ \left(1 - \sum_{j=1}^{p} \alpha_j\right) \qquad Q\ E\ D$$

PROOF OF THEOREM 3 Clearly, under the conditions, $h(\xi_t) \geq \alpha_0 > 0$, establishing part (a) Let

$$\phi_{i,m,t} = E(|\partial h(\xi_t)/\partial \alpha_i|\partial h(\xi_t)/\partial \xi_{t-m} \mid \psi_{t-m-1})$$

$$= 2\alpha_m E(|\xi_{t-i}|^2 |\xi_{t-m}| \psi_{t-m-1})$$

Now there are three cases; $i > m$, $i = m$, and $i < m$ If $i > m$, then $\xi_{t-i} \in \psi_{t-m-1}$ and the conditional expectation of $|\xi_{t-m}|$ is finite, because the conditional density is normal If $i = m$, then the expectation becomes $E(|\xi_{t-m}|^3 \mid \psi_{t-m-1})$ Again, because the conditional density is normal, all moments exist including the expectation of the third power of the absolute value If $i < m$, the expectation is taken in two parts, first with respect to $t - i - 1$

$$\phi_{i m t} = 2\alpha_m E\left\{|\xi_{t-m}| E(\xi_{t-i}^2 \mid \psi_{t-i-1}) \mid \psi_{t-m-1}\right\}$$

$$= 2\alpha_m E\left\{|\xi_{t-m}|\alpha_0 + \sum_{j=1}^{p} \alpha_j \xi_{t-i-j}^2\Big) \mid \psi_{t-m-1}\right\}$$

$$= 2\alpha_m \alpha_0 E\left(\xi_{t-m} \mid \psi_{t-m-1}\right) + \sum_{j=1}^{p} \alpha_j \phi_{i+j\, m\, t}$$

In the final expression, the initial index on ϕ is larger and, therefore, may fall into either of the preceding cases, which, therefore, establishes the existence of the term. If there remain terms with $i + j < m$, the recursion can be repeated As all lags are finite, an expression for $\phi_{i m t}$ can be written as a constant times the third absolute moment of ξ_{t-m} conditional on ψ_{t-m-1}, plus another constant times the first absolute moment As these are both conditionally normal, and as the constants must be finite as they have a finite number of terms, the second part of the regularity condition has been established *Q.E.D.*

To establish Theorem 4, a careful symmetry argument is required, beginning with the following lemma.

LEMMA: *Let u and v be any two random variables $E(g(u, v) \mid v)$ will be an anti-symmetric function of v if g is anti-symmetric in v, the conditional density of $u \mid v$ is symmetric in v, and the expectation exists.*

PROOF:

$$E(g(u, -v) \mid -v) = -E(g(u, v) \mid -v) \qquad \text{because } g \text{ is anti-symmetric in } v$$

$$= -E(g(u, v) \mid v) \qquad \text{because the conditional density is symmetric.}$$

Q E D

PROOF OF THEOREM 4 The i, j element of $I_{\alpha\beta}$ is given by

$$(I_{\alpha\beta})_{ij} = \frac{1}{2T}\sum_t E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \beta_j}\right)$$

$$= -\frac{1}{2T}\sum_t \sum_{m=1}^{p} E\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}} x_{jt-m}\right] \qquad \text{by the chain rule}$$

If the expectation of the term in square brackets, conditional on ψ_{t-m-1}, is zero for all i, j, t, m, then the theorem is proven

$$E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}} x_{jt-m} \mid \psi_{t-m-1}\right) = x_{jt-m} E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}} \mid \psi_{t-m-1}\right)$$

because x_{jt-m} is either exogenous or it is a lagged dependent variable, in which case it is included in ψ_{t-m-1}

$$\left|E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}} \mid \psi_{t-m-1}\right)\right| \le E\left(\frac{1}{h_t^2}\left|\frac{\partial h_t}{\partial \alpha_i}\right|\left|\frac{\partial h_t}{\partial \epsilon_{t-m}}\right| \mid \psi_{t-m-1}\right)$$

$$\le \frac{1}{\delta^2} E\left(\left|\frac{\partial h_t}{\partial \alpha_i}\right|\left|\frac{\partial h_t}{\partial \epsilon_{t-m}}\right| \mid \psi_{t-m-1}\right)$$

by part (a) of the regularity conditions and this integral is finite by part (b) of the condition. Hence, each term is finite. Now take the expectation in two steps, first with respect to ψ_{t-m}. This must therefore also be finite

$$E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}}\mid \psi_{t-m}\right) = g(\epsilon_{t-m}).$$

By the symmetry assumption, h_t^{-1} is symmetric in ϵ_{t-m}, $\partial h_t/\partial \epsilon_{t-m}$ is anti-symmetric. Therefore, the whole expression is anti-symmetric in ϵ_{t-m}, which is part of the conditioning set ψ_{t-m}. Because h is symmetric, the conditional density must be symmetric in ϵ_{t-m} and the lemma can be invoked to show that $g(\epsilon_{t-m})$ is anti-symmetric.

Finally, taking expectations of g conditional on ψ_{t-m-1} gives zero, because the density of ϵ_{t-m} conditional on the past is a symmetric (normal) density and the theorem is established. *Q E D*

REFERENCES

[1] AMEMIYA, T. "Regression Analysis when the Variance of the Dependent Variable is Proportional to the Square of Its Expectation," *Journal of the American Statistical Association*, 68(1973), 928–934.

[2] ANDERSON, T. W.. *The Statistical Analysis of Time Series*. New York: John Wiley and Sons, 1958

[3] BELSLEY, DAVID "On the Efficient Computation of Non-Linear Full-Information Maximum Likelihood Estimator," paper presented to the European Meetings of the Econometric Society, Athens, 1979.

[4] BREUSCH, T S, AND A R PAGAN: "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 46(1978), 1287–1294.

[5] ——— "The Lagrange Multiplier Test and Its Applications to Model Specification," *Review of Economic Studies*, 47(1980), 239–254

[6] COX, D R, AND D V HINKLEY: *Theoretical Statistics* London: Chapman and Hall, 1974

[7] CROWDER, M. J "Maximum Likelihood Estimation for Dependent Observations," *Journal of the Royal Statistical Society*, Series B, 38(1976), 45–53

[8] DAVIDSON, J E H, D. F HENDRY, F. SRBA, AND S. YEO "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *The Economic Journal*, 88(1978), 661–691

[9] ENGLE, R F "A General Approach to the Construction of Model Diagnostics Based upon the Lagrange Multiplier Principle," University of California, San Diego Discussion Paper 79-43, 1979.

[10] ——— "Estimates of the Variance of U S Inflation Based on the ARCH Model," University of California, San Diego Discussion Paper 80-14, 1980

[11] FRIEDMAN, MILTON: "Nobel Lecture: Inflation and Unemployment," *Journal of Political Economy*, 85(1977), 451–472.

[12] GODFREY, L G.: "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46(1978), 1293–1302

[13] GRANGER, C W J, AND A ANDERSEN *An Introduction to Bilinear Time-Series Models*. Göttingen. Vandenhoeck and Ruprecht, 1978.

[14] KHAN, M. S "The Variability of Expectations in Hyperinflations," *Journal of Political Economy*, 85(1977), 817–827

[15] KLEIN, B. "The Demand for Quality-Adjusted Cash Balances: Price Uncertainty in the U.S Demand for Money Function," *Journal of Political Economy*, 85(1977), 692–715

[16] LUCAS, R. E, JR: "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, 63(1973), 326–334

[17] MCNEES, S. S.. "The Forecasting Record for the 1970's," *New England Economic Review*, September/October 1979, 33–53

[18] WHITE, H: "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, 48(1980), 817–838.

Econometrica, Vol. 55, No. 2 (March, 1987), 391-407

ESTIMATING TIME VARYING RISK PREMIA IN THE TERM STRUCTURE: THE ARCH-M MODEL[1]

BY ROBERT F. ENGLE, DAVID M. LILIEN, AND RUSSELL P. ROBINS

The expectation of the excess holding yield on a long bond is postulated to depend upon its conditional variance. Engle's (1982a) ARCH model is extended to allow the conditional variance to be a determinant of the mean and is called ARCH-M. Estimation and inference procedures are proposed and the model is applied to three interest rate data sets. In most cases the ARCH process and the time varying risk premium are highly significant. A collection of LM diagnostic tests reveals the robustness of the model to various specification changes such as alternative volatility or ARCH measures, regime changes, and interest rate formulations. The model explains and interprets the recent econometric failures of the expectations hypothesis of the term structure.

KEYWORDS: Term structure, financial models, ARCH, risk premium, heteroskedasticity, nonlinear models.

1. INTRODUCTION

ALTHOUGH THE VALUATION of risk is the central feature of financial economics, the standard methods for measuring and predicting risk are extraordinarily simple and unsuited for time series analysis. As the degree of uncertainty in asset returns varies over time, the compensation required by risk averse economic agents for holding these assets, must also be varying. Time series models of asset prices must therefore both measure risk and its movement over time, and include it as a determinant of price. Any increase in the expected rate of return of an asset as it becomes more risky will be identified as a risk premium.

The importance of such risk premia in the term structure of interest rates has been highlighted by a series of papers which all find the traditional expectations hypothesis inadequate to explain the observed data. For some recent examples see Shiller (1979, 1981), Sargent (1979, 1972), Shiller, Campbell, and Schoenholtz (1983), Mankiw and Summers (1984), and Campbell (1984). Some of these are based upon tests which find the variance of long term rates too large to be consistent with the expectations hypothesis. Others are based on regression tests which essentially show that the implicit predictors of future interest rates, derivable from the term structure, are inefficient and biased. Information available at the time could have improved the accuracy of the forecasts. Stated another way, these tests find that the one period rate of return which should, ex ante, be unforecastable, could have been predicted using available information.

These findings are generally interpreted as implying either some form of less than fully rational expectations, or time varying premia on different term debt. Attempts by Shiller, Campbell, and Schoenholtz (1983) and Mankiw and Summers (1984) to model particular forms of irrational expectations were unsuccessful.

[1] The authors are indebted to many for helpful comments including Bob Shiller, Larry Summers, Clive Granger, Ross Starr, Ken Wallis, David Hendry, Larry Weiss, and James Tobin, but retain responsibility for remaining errors. Computations were carefully carried out by Tim Bolleralev and Yoahi Baba.

Consequently, the main thrust of this literature is to introduce the possibility of time varying term premia. Amsler (1984) and Pesando (1983) have extended Shiller's variance bounds to allow time varying term premia. Campbell (1984) and Mankiw and Summers (1984) estimate or derive statistics about the required properties of time varying term premia. The latter conclude: "Most of the changes in the slope of the yield curve reflect these changing liquidity premiums or expectations that do not satisfy the standard postulates of rationality. These results suggest the importance of developing models capable of explaining fluctuating liquidity premiums."

The key postulate in the current paper is that time varying premia on different term debt instruments can be well modeled as risk premia where the risk is due to unanticipated interest rate movements and is measured by the conditional variance of the one period holding yield. While this is in the spirit of Bodie, Kane, and McDonald (1983) and Fama (1976), new econometric techniques are needed to estimate and test this model and these are developed here.

The autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982a), explicitly models time varying conditional variances by relating them to variables known from previous periods. In its standard form the ARCH model expresses the conditional variance as a linear function of past squared innovations; in markets where price is a Martingale, price changes are innovations, and this corresponds precisely to the Mandlebroit (1963) observation: "Large changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes . . ." The ARCH model is used to provide a rich class of possible parameterizations of heteroscedasticity.

This paper introduces the ARCH-M model which extends the ARCH model to allow the conditional variance to affect the mean. In this way changing conditional variances directly affect the expected return on a portfolio. This resolves many of the empirical paradoxes in the term structure. Variables which apparently were useful in forecasting excess returns are correlated with the risk premia and lose their significance when a function of the conditional variance is included as a regressor. Furthermore, the heteroscedasticity in the disturbances had biased the test statistics, leading to the false finding of significant variables.

This model is applied to six month treasury bills, to two month treasury bills, and to 20 years Aaa corporate bonds to determine whether there appear to be time varying risk premia and how large they are. Section 2 develops a theoretical model of the relationship between means and variances which is formulated as a statistical model in Section 3. Section 4 describes the ARCH-M model and Sections 5 and 6 present the applications. Section 7 is a summary.

2. A MODEL OF THE RELATION BETWEEN RISK AND RETURN

Risk averse economic agents require compensation for holding risky assets. In the simplest set-up of one risky asset with normally distributed returns and one riskless asset, the risk is measured by the variance of the returns from holding the asset, and the compensation by a rise in the expectation of the return. The relation between the mean and the variance of the returns which will insure that the asset is fully held in equilibrium will depend upon the utility function of the agents and the supply conditions of the assets.

To investigate this relation we now suppose that in this two asset economy the variance of the payoff of the risky asset may change over time and consequently the price offered by risk averse agents will change over time. This equilibrium price determines the relation between the mean and variance of the excess returns from holding the risky asset and therefore how the risk premium is related to the variance of the returns.

Consider a world with two assets, one has price 1 and is perfectly elastically supplied at a sure total rate of return r. The other has a price p and yields a random total return q (denominated in units of the numeraire) which has mean θ and variance ϕ. Wealth W, measured in units of the riskless asset, is therefore allocated between shares of the sure asset x, and shares of the risky asset s, so that

(1) $$W = ps + x.$$

The excess return per dollar invested in shares of the risky asset is given by

$$y = (q/p) - r,$$

so that the mean and variance of the excess returns is given by

(2) $$E(y) = \mu = (\theta/p) - r, \qquad V(y) = \sigma^2 = \phi/p^2.$$

Agents maximize expected utility of the end-of-period wealth, which, assuming normality of the returns, means that only the first two moments of the distribution matter. Under constant absolute risk aversion, expected utility can be expressed by:

$$EU = 2E(qs + rx) - bV(qs + rx)$$

and it will be maximized by choosing

(3) $$sp = \mu/(b\sigma^2).$$

Now suppose ϕ has a time subscript and is known to agents although not to the econometrician. Then the equilibrium values of p, μ, σ^2, and s will also vary over time. If in equilibrium the *value* of the outstanding shares of the risky asset remains constant, then the mean return will be proportional to the variance of returns since $s_t p_t$ in (3) is a constant.

A convenient assumption is that the riskless asset is held in zero net supply so that r becomes endogenous. The value of the outstanding shares of the risky asset is simply W. The mean and variance will therefore be proportional regardless of the supply elasticity of s if both wealth and b are constant. Such a model, however, leaves no role for price in evaluating risk.

If, instead, the physical number of shares is fixed so that $s_t = s$ and r is fixed, then in equilibrium (4) can be rewritten

$$\mu_t^2 + \mu_t r_t = bs\sigma_t^2\theta$$

and, suppressing time subscripts,

(4) $\mu = [-r + \sqrt{r^2 + 4bs\sigma^2\theta}]/2$

so that the mean will be zero when the variance is zero, the slope is always positive, and for large variance the mean is proportional to the standard deviation. Thus if ϕ varies over time, but r, s, and θ do not, the econometrician should expect to see a relation between observed means and variances of returns which makes them move in the same direction but not proportionally.

For more general utility functions b will itself be a function of other variables such as σ^2. Thus we can replace b in (4) with $b(\sigma^2)$. Furthermore, there may be some elasticity of supply of the risky asset so that

$$s = f(p) = f(\theta/(\mu + r))$$

can be substituted for s. With these two flexible functions it is possible to find a wide range of relationships between observed means and variances.

Thus in general, one might expect the mean to increase less than in proportion to the variance with the precise relation determined by the supply elasticity of the risky (and possibly the riskless) asset and the risk preferences of agents. This paper introduces some empirical evidence on this relationship.

3. FORMULATION OF THE MODEL

Letting μ_t be the risk premium, y_t the excess holding yield on a long bond relative to a one period treasury bill, and ε_t the difference between the ex ante and ex post rate of return which in efficient markets would be unforecastable,

(5) $y_t = \mu_t + \varepsilon_t, \qquad \mathrm{Var}\,(\varepsilon_t \,|\, \text{all available information}) = h_t^2.$

It is assumed that the risk in holding a long bond is not diversifiable so that only the variance matters. The initial specification takes the mean as a linear function of the standard deviation:

(6) $\theta_t = \beta + \delta h_t.$

A nonzero value of β might reflect the linearization of a nonlinear function such as that derived above, or a preferred habitat argument. The choice of the standard deviation represents the assumption that changes in variance are reflected less than proportionally in the mean. Empirically, the log of h_t is found to be even better.

A complication in the interpretation of θ_t arises from the differential tax treatment of capital gains and interest income. Under the tax laws, long term capital gains are taxed at a lower rate than ordinary interest income and short term capital gains. This feature of the tax system makes a strategy of investing in long term bonds more desirable than rolling over short term paper. Investors can, to a large extent, treat one period capital losses as ordinary income for tax purposes by selling the bond and realizing their losses. Short term capital gains can be turned into long term capital gains for tax purposes by holding the bond for a year or longer. Because this choice can be made ex post, after Y_t is observable, risk neutral investors should be willing to hold long term bonds at a lower expected pre tax yield than is paid on treasury bills. This tax advantage may explain the fact that the average value of Y_t for many types of long term bonds, has been below the average short term treasury bill rate over the last 30 years. We might therefore expect $\beta < 0$.

To complete the specification of the model, h_t^2, the conditional variance, must be parameterized as a function of the information set available to investors. We assume that the most useful information to agents are the previous innovations or surprises ε_t. If these have been large in absolute value then, extending Mandlebroit's observation, they are likely to be large in the future. In its simplest form we postulate that

$$(7) \qquad h_t^2 = \alpha_0 + \alpha_1 \sum_{i=1}^{p} w_i \varepsilon^2_{t-i}.$$

The conditional variance as observed by both the economic agents and the econometrician is a weighted sum of past squared surprises. One can discount older innovations in this weighting scheme.

Other variables which are in the information set at time t could also be introduced into (7) in the fashion of more traditional heteroscedasticity corrections. One such suggestion would be to use the squared changes in price as analyzed by Mandlebroit. Such a specification misses the fact that in the bond market a portion of the price change may be anticipated and this information is unlikely to be useful in forecasting changes in variance.

In the next section, the estimation and testing of the model in (5), (6), and (7) is considered in a more general context. In the following three empirical analyses, many of the caveats discussed above are then put to test.

4. ESTIMATING AND TESTING THE ARCH-M MODEL

The economic model described in the previous section incorporates an important extension of Engle's (1982a) ARCH model or in fact any heteroscedastic model; not only are the disturbances heteroscedastic, but the standard deviation of each observation affects the mean of that observation. In this section the estimation and testing of such models, called ARCH in mean or ARCH-M models, is discussed.

The general setup is given by

$$(8) \qquad Y_t \mid X_t, \Pi_t \sim N(\beta' X_t + \delta h_t, h_t^2),$$

$$(9) \qquad h_t^2 = \alpha' W_{\eta t} + \gamma' Z_t,$$

where X_t and Z_t are $k \times 1$ and $j \times 1$ vectors of weakly exogenous and lagged dependent variables, as in Engle, Hendry, and Richard (1983). The vector Z_t includes a constant whose coefficient represents the constant variance component of h_t. The $p \times 1$ vector $\eta_t' = (\varepsilon_{t-1}^2, \ldots, \varepsilon_{t-p}^2)$ where ε_t are the disturbances given by $Y_t - \beta' X_t - \delta h_t$. The matrix W is a $q \times p$ array of fixed constants which may be used to impose restricted parameterizations on the response of the conditional variance to past squared residuals. In the most unrestricted case, W would be the identity matrix. The variance parameter vectors α and γ are therefore $q \times 1$ and $j \times 1$ respectively while the mean parameter vectors β and δ are $k \times 1$ and 1×1. These parameters can be combined into $\phi' = (\alpha', \gamma', \beta', \delta)$, an $m \times 1$ vector where $m = q + j + k + 1$.

Conditional on the initial values of all the data, the log likelihood function can be expressed as

$$(10) \qquad L(\phi) = \sum_t L_t(\phi); \quad L_t(\phi) = -\log h_t - \varepsilon_t^2 / 2h_t^2.$$

In practice, the presample values of the disturbances are set to their expectation, zero. The first order conditions for a maximum of this likelihood are given by:

$$(11) \qquad \partial L_t / \partial \phi = \sum ([\varepsilon_t^2 - h_t^2 - h_t \delta \varepsilon_t] h_t^{-4}) \, \partial h_t^2 / \partial \phi / 2 \\ - \sum [\varepsilon_t / h_t^2][\partial \beta' / \partial \phi$$

The derivatives of the parameters with respect to ϕ are simply matrices with zeros and ones which select which terms to include for each derivative. The second line of (11) is the term relevant for GLS estimation of the regression coefficients without ARCH complications, that is when $\alpha = 0$. The expression in (11) gives the standard ARCH model when δ is zero.

The primary complexity introduced in this model comes in evaluating $\partial h^2 / \partial \phi$. From (9) this depends upon the derivatives of previous innovations with respect to the parameters. Yet these derivatives in turn depend upon the past derivatives of h with respect to the parameters if δ is nonzero. The desired derivatives must be computed recursively from an assumption that the initial values do not depend upon the parameters.

In the early analyses presented in Engle, Lilien, and Robins (1982) summarized in Section 5, analytical derivatives were calculated recursively and used to evaluate (11). However, numerical derivatives gave similar results, were simpler to compute and gave added flexibility to changes in specification. They therefore are probably the preferred approach for the ARCH-M model.

Estimation and testing can simply be carried out in terms of these derivatives. $\partial L / \partial \phi$ can be written compactly in terms of the $T \times m$ array S with typical element

$$[S]_{ti} = \partial L_t / \partial \phi_i$$

as

$$(12) \qquad \partial L / \partial \phi = S' i$$

where i is a $T \times 1$ unit vector so the first order condition is simply

$$S' i = 0.$$

The Hessian of the log likelihood is the sum of the Hessians of the t conditional log likelihoods, L_t. Under the assumption that the likelihood function is correctly specified,

$$\mathcal{I}_t = E[\partial L_t/\partial\phi \; \partial L_t/\partial\phi'] = -E[\partial^2 L_t/\partial\phi \; \partial\phi']$$

where $\mathcal{I}_t$ is the information matrix of the tth observation. Defining the information in the sample $\mathcal{I}$ is the average of the information over each observation,

$$\mathcal{I} = E[S'S/T].$$

Under slightly stronger conditions, $S'S/T$ is also consistent for $\mathcal{I}$.

A ready solution to the maximization of this likelihood function is to adopt the Berndt, Hall, Hall, Hausman (1974) approach using the iteration

$$\phi^{i+1} = \phi^i + \lambda(S'S)^{-1}S'i \tag{13}$$

with λ as a step length which is adjusted from its a priori value of unity by a simple line search, and S as the matrix of first derivatives evaluated at ϕ^i.

The likelihood is in the form analyzed by Crowder (1976). Under sufficient regularity conditions, a solution to (13) will have the property that

$$(S'S)^{1/2}(\phi^* - \phi^0) \overset{A}{\sim} N(0, I) \tag{14}$$

where ϕ^* is the maximum likelihood estimator obtained from (13) and ϕ^0 is the true value of the parameters. Unlike the simple ARCH model, this information matrix is not block diagonal between the parameters of the mean and the parameters of the variance.

Pantula (1984) has carefully investigated regularity conditions sufficient to guarantee (14) in the simple first order ARCH case. His conditions are stronger than can be accepted for this study in that he requires the existence of eighth order moments of the disturbance which are only finite for very small values of the ARCH parameter. Weiss (1986) has suggested some slightly weaker conditions; however, neither has addressed the ARCH-M model. Thus the appropriateness of the asymptotic distribution theory for this analysis remains a conjecture at this point.

Subject to the above caveat, inference procedures are available directly from (14). In particular, Wald tests can be computed in standard fashion. Lagrange multiplier tests can be simpler if the model has already been estimated under the null hypothesis and are easily constructed from the matrix of scores, S. Suppose the null hypothesis specifies that $\phi \in \Phi^0$ which is a proper subset of Φ. Denote by S^0 the matrix of scores calculated assuming the more general model to be true, but evaluated at the parameter estimates under the null. The scores corresponding to the restricted parameters are the Lagrange multipliers, and their variances are given by the information matrix. The LM test can be constructed as

$$\begin{aligned} \Phi_{LM} &= i'S^0(S^{0\prime}S^0)^{-1}S^{0\prime}i \\ &= TR_0^2 \end{aligned} \tag{15}$$

where R_0^2 is the uncentered R^2 achieved by regressing the unit vector on the matrix of scores under the null. This statistic will asymptotically be chi squared with the number of degrees of freedom of the restriction when the null is true. This is easily computed from the R^2 of the first iteration of (13) starting from the estimates found under the null. Thus the tests take a form familiar from Engle (1982b, 1984) and it is recommended to construct a battery of diagnostics to convey information on the validity of the model both to the user and the reader.

The LM tests are convenient for testing restrictions in either the mean or the variance specification since reestimation may be costly and convergence is sometimes unsure. Tests are easily constructed for variables excluded from the mean such as interest rates or other functional forms. It is just as simple to test variance restrictions such as $\alpha = 0$, α is a set of linearly declining weights, or elements of γ are equal to zero (thereby testing for variables excluded from h). Many of the variance tests, however, may be interpreted as being on the boundary of the admissible parameter space so that one-tailed tests or other adjustments may be appropriate.

For the preferred models in this study h_t depended only on the intercept and a weighted average of past squared innovations where the weights are assumed to be linearly declining. These strong restrictions are subjected to a great variety of tests which allow changes in slope, seasonal spikes, freely estimated coefficients, and a wide variety of observable variables such as interest rates, volatility, and dummy variables for policy regimes. The models generally accept the more parsimonious specification at reasonable significance levels either because they are close to the true specification or because there is little power in the data to discriminate between alternative variance formulations. If the models with less restricted parameterizations are iterated toward convergence (for example to calculate a Wald or a likelihood ratio test) we found it difficult to prevent nonnegativities in the parameters regardless of the types of penalty functions or transformations considered. In this case there were likely to be many local maxima and generally the likelihood was ill-behaved. Thus the imposition of a parsimonious specification for the variance function such as linearly declining weights appears to be statistically supportable, computationally useful, and economically sensible.

5. THE RESULTS FOR SHORT TERM T-BILLS

Using Salomon Brothers data from the Analytical Record of Yields from 1960 through 1984 II on 3 and 6 month treasury bills, the excess holding yield, y_t, was calculated as:

$$y_t = [(1+R_t)^2/(1+r_{t+1})] - (1+r_t)$$

which is approximately

$$y_t \approx 2R_t - r_{t+1} - r_t$$

where R_t is the yield on a six month bill and r_t is the three month yield, each measured at the beginning of the quarter.

Regressing the excess holding yield on a constant gives

$$(16) \qquad \underset{(4.04)}{y_t = .142} + e_t, \qquad s = .351,$$

$$L = 51.1.$$

Thus, the mean of the excess holding yield over the sample period is .142 per cent at quarterly rates or .568 per cent at annual rates. The standard deviation is .35 at quarterly rates. From the linearized expression for the excess holding yield above, the average yield spread was half .568 per cent or .284 per cent at annual rates. The maximum return on a three month balanced portfolio obtained by borrowing at the three month rate and lending at the six, was 8.2 per cent at annual rates. The worst return occurred in the subsequent quarter and was −3.1 per cent. The rates of return from such portfolios are quite erratic and, as expected, are not large especially if transaction costs are important in forming these portfolios.

A glance at the solid line in Figure 1 confirms the changes in variance which are hypothesized by the ARCH-M model to account for the changing risk premia. The vertical axis is measured in quarterly percentage rates of return. Clearly, the period subsequent to the 1979 change in operating procedures shows substantially more variability than earlier periods; however, there are also earlier episodes of increased variability. Regressing the squared residuals on a fourth order linearly declining weighted average of past squared residuals gives the ARCH test as $TR^2 = 10.1$ which would be X_1^2 if there were no ARCH. There is clearly strong evidence of heteroscedasticity in the errors.

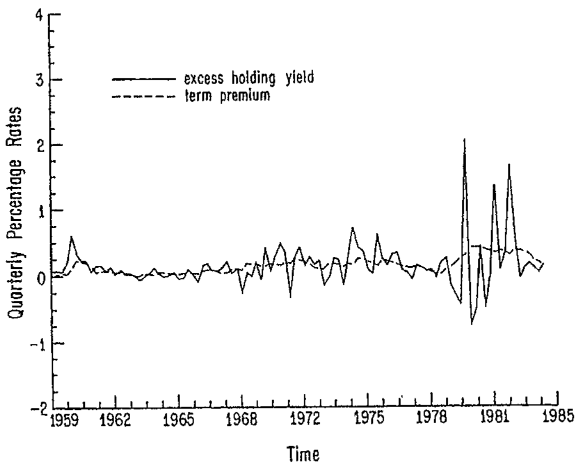

FIGURE 1—Excess hold yield of 6 month Treasury Bills and estimated risk premia.

Regressing the excess holding yield on a constant and allowing ARCH disturbances of fourth order gives:

$$(17)\qquad \underset{(3.77)}{y_t = .048} + e_t, \qquad h_t^2 = \underset{(.95)}{.004} + \underset{(7.3)}{1.90} \sum_{\tau=1,4} w_\tau \varepsilon_{t-\tau}^2$$

$$L = 85.17, \qquad w_\tau = (5-\tau)/10 \qquad (\tau = 1, \ldots, 4).$$

The ARCH effect is very strong, showing a t statistic of 7.3. The magnitude is also very large as values over 1 imply nonstationary variance processes. The estimate of the mean changes dramatically when the high variance periods are given less weight in the regression; the constant term premium falls to .048 per cent at a quarterly rate or .2 per cent at annual rates.

The time varying risk premium has been swept into the disturbance term in (17) and represents misspecification. The hypothesized true model, as presented in Section 2, can be formalized as:

$$(18)\qquad y_t = \beta + \delta h_t + \varepsilon_t,$$

$$\varepsilon_t/\text{past information} \sim N(0, h_t^2),$$

$$h_t^2 = \gamma + \alpha \sum_{\tau=1,4} w_\tau \varepsilon_{t-\tau}^2, \qquad w_\tau = (5-\tau)/10 \qquad (\tau = 1, \ldots, 4).$$

The maximum likelihood estimates and their t statistics are:

$$(19)\qquad y_t = \underset{(-1.29)}{-.0241} + \underset{(5.15)}{.687}\, h_t + e_t,$$

$$h_t^2 = \underset{(1.08)}{.0023} + \underset{(6.30)}{1.64} \sum w_\tau \varepsilon_{t-\tau}^2,$$

$$L = 96.34, \qquad w_\tau = (5-\tau)/10 \qquad (\tau = 1, \ldots, 4).$$

As can be easily seen, all the slope coefficients are highly significant, indicating that there is not only an ARCH effect ($\alpha \neq 0$), but also a time varying risk premium ($\delta \neq 0$). The expected riskless return is negative but not significantly so and the minimum possible expected return which would be achieved if all recent forecasts had been precisely correct, is very small and positive (.0009). The risk premium is two thirds of the standard deviation of the return, which is quite substantial, indicating stronger risk aversion by the borrowers than the lenders in this market.

The parameter in the ARCH equation is above one which implies that the unconditional variance of the excess holding yield is infinite with a fat tailed distribution. The conditional distribution,which for most purposes is the relevant distribution, is of course still normal with a finite variance. An arbitrarily large return could occur if a sufficiently long string of innovations were all large. Such an episode would be easily reversed by a number of innovations near their median value of zero. Simulations of this situation show rather sensibly behaved series with larger bursts of volatility than would be expected from a marginally normal random variable. It is possible that the maximum likelihood estimates will not have their standard properties, but, as in the unit root case, they may have superior convergence rates and correctly calculated standard errors. As mentioned in the previous section, the asymptotic distribution theory for this problem remains to be solved. The infinite unconditional variance may be related to the frequent failures of the variance bounds tests for interest rates.

A series of diagnostic tests were calculated for the model in (19). Although several were significant, the tests for the functional relationship between the risk and rate of return are of particular interest. LM tests for omitted variables h_t^2, $\exp(h_t)$, and $\log(h_t)$ were computed to test the assumed linearity between the standard deviation and mean of returns. Economic theory has little to say on the nature of this trade-off as it presumably depends on the risk preferences of the traders. Only the log variable was significant with a test statistic of 4.13. Estimating the model with both h_t and $\log(h_t)$ produced t statistics of 2.0 on the log and $-.4$ on the level and a log likekihood of $L = 101.62$, thereby confirming that the model with the log of standard deviation is superior to that in the level of the standard deviation.

The final preferred model is therefore:

$$\begin{aligned} (20)\qquad & y_t = \underset{(4.38)}{.355} + \underset{(3.36)}{.135} \log h_t + e_t, \\ & h_t^2 = \underset{(2.22)}{.005} + \underset{(5.56)}{1.48} \sum_{\tau=1,4} w_\tau \varepsilon_{t-\tau}^2, \\ & L = 101.35, \qquad w_\tau = (5-\tau)/10. \end{aligned}$$

In this model all the coefficients are significant and the log likelihood is substantially above that of (19). The minimum term premium occurring when all past innovations are zero is now a very small negative value of $-.008$ per cent at quarterly rates.

Several sets of diagnostic tests were performed with this model as well. These are summarized in Table I. Volatility is defined by:

$$\text{Volatility} = \sum_{\tau=1,4} w_\tau y_{t-\tau}^2, \qquad w_\tau = (5-\tau)/10,$$

so that it differs from the ARCH variance by the time varying risk premium. One would expect that the weighted average of residuals would give a better estimate of the true residual variance than the same function of the dependent variable; however there is no guarantee. Table I shows the robustness of the model in (20) to a variety of types of misspecification. None of the tests is significant at the 5 per cent level. The tests check for nonlinearities in the risk premium, volatility, structural shifts in October 1979, and misspecifications of the ARCH process through omitted variables or inappropriately applied constraints. The ARCH model with log Volatility alone achieves only log likelihood $L = 98.4$ although the significance and size of the variables is nearly the same as in (20).

The economically most interesting test is that for the yield spread and we turn to a more careful analysis of this model. Mankiw and Summers (1984) (MS) find that the yield spread is a significant and positive determinant of the excess holding

TABLE I
DIAGNOSTIC TESTS FOR ARCH-M MODEL (20)

Variable	TR^2		Distribution
Variables Omitted from the Mean			
h_t	.31	~	χ_1^2
h_t^2	1.67	~	χ_1^2
Volatility	1.44	~	χ_1^2
Log Volatility	.50	~	χ_1^2
Post October 1979 Dummy	.38	~	χ_1^2
r_t	.60	~	χ_1^2
R_t	.83	~	χ_1^2
$R_t - r_t$	2.92	~	χ_1^2
y_{t-1}	.14	~	χ_1^2
y_{t-4}	3.38	~	χ_1^2
Variables Omitted from the Variance			
Volatility	.27	~	χ_1^2
Post October 1979 Dummy	.07	~	χ_1^2
r_t	1.64	~	χ_1^2
R_t	1.60	~	χ_1^2
$R_t - r_t$	.90	~	χ_1^2
ε_{t-1}^2	.31	~	χ_1^2
ε_{t-4}^2	.62	~	χ_1^2
$\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \varepsilon_{t-3}^2$	3.11	~	χ_3^2
$\sum w_\tau \varepsilon_{t-\tau}^2, w_\tau = (13-\tau)/78, \tau = 1, \ldots, 12$	.76	~	χ_1^2

yield. This implies a failure in the expectations hypothesis and a failure of an alternative hypothesis that long rates are overly sensitive to short rates. Our data set gives the following least squares estimate for this model:

$$y_t = -.50 + 2.44\,(R_t - r_t) + e_t, \qquad \sigma = .312. \tag{21}$$
$$(-1.10)\,(5.46)$$

The corresponding coefficient and t statistic in MS for the yield spread are 1.72 and 3.1 respectively. Their data set is a little shorter, from a different source and embodies the Shiller linearizations.

Adding the yield spread to model (20) gives:

$$y_t = .325 + .130 \log h_t + .392\,(R_t - r_t) + e_t, \tag{22}$$
$$(4.28)\,(3.59) \qquad (2.58)$$
$$h_t = .004 + 1.64 \sum w_\tau \varepsilon_{t-\tau}^2,$$
$$(1.38)\ \ (4.86)$$
$$L = 103.48, \qquad w_\tau = (5-\tau)/10 \qquad (\tau = 1, \ldots, 4).$$

It now can be seen that by both Wald and LR tests the yield spread is a significant determinant at the 5 per cent but not 1 per cent level and by the LM test it is significant at the 10 per cent but not 5 per cent level. By economic standards the size of the coefficient on the yield spread has fallen dramatically from the least

TABLE II
ESTIMATES OF VARIOUS ARCH-M MODELS
EXCESS HOLDING YIELD OF 6 MONTH T-BILLS

Indep	59.1-84 2	59.1-71 3	71.4-84 2	59.1-79.3	61.3-74.1
Log h_t	.135 (3.36)	.092 (3.88)	.196 (2.40)	.177 (2.96)	.093 (2.01)
Const.	.355 (4.38)	.272 (4.31)	.455 (3.36)	.446 (3.72)	.261 (2.52)
ARCH α	1.48 (5.56)	1.67 (5.15)	1.49 (3.57)	1.25 (4.60)	1.20 (2.84)

squares fit. The rest of the parameter estimates are very close to those obtained before in (20). Economically, it is not surprising to find some residual effect in the yield spread. The expected value of the spread is approximately proportional to the risk premium this period. Since it is highly autocorrelated, it will be a very good predictor of the risk premium next period. If information other than past innovations is useful in forecasting risk premia, then one might expect to find a significant coefficient on the past yield spread. A useful extension would be to allow the yield spread to directly influence the variance and consequently to indirectly influence the risk premium.

As much of the variance in interest rates is concentrated at the end of the sample period, the model was reestimated using subsets of the data. Surprisingly, the results are relatively insensitive to the sample period both in magnitude and in significance. See Table II.

Figure 1 plots the excess holding yield and the estimated risk premium. The scale is in quarterly percentage rates of return. The term premium rises to its highest value (.41 per cent quarterly or 1.64 per cent annual rates) in the fourth quarter of 1980. Over the sample period there are two values which are very slightly negative. On average, the term premium is .14 per cent. Although the most interesting and noticeable rise in the term premium is 1979–1984, there are also relative increases in 1960, 1972, and 1975, each of which is accompanied by an increase in volatility of the excess holding yield.

6. MODELLING OTHER INTEREST RATES

Two additional interest rate series have been modelled using the ARCH-M model and more are in progress. The first is the monthly data set constructed by Fama (1976) on two month vs. one month treasury bills from 1953.1 to 1971.7. The data set differs from that used above in the sampling interval and in the sample period. In this case the holding period is naturally taken to be one month rather than one quarter and consequently the riskless asset is the one month treasury bill rather than two or three month treasury bills. If a quarter is the correct interval, then shorter lived assets must be rolled over at uncertain rates and therefore, the short term asset would be the risky one. For a theoretical discussion of these issues see Woodward (1983).

The model in (18) was estimated directly although a longer lag was allowed in the ARCH process to give a comparable memory to the variance estimator. The results are:

$$(23)\qquad \underset{}{y_t} = \underset{(-1.2)}{-.00052} + \underset{(4.7)}{.80}\, h_t, \qquad h_t^2 = c_0 + \underset{(8.6)}{1.13} \sum_{\tau=1,12} w_\tau \varepsilon_{t-\tau}^2,$$

$$w_\tau = (13-\tau)/78 \qquad (\tau = 1, \ldots, 12).$$

These are quite similar to those in equation (19) where in both cases the ARCH parameters are in the explosive range and the coefficient of the standard deviation is highly significant with a value of .69 before and .8 here. The estimated risk premium is plotted in Figure 2.

A somewhat different result was obtained using 20 year AAA corporate bonds from 1953.1 to 1980.2. Assuming that the bonds are effectively infinitely lived, the one quarter excess holding yield can be expressed in terms of the quarterly yield to maturity, R_t, and the three month treasury bill rate, r_t:

$$y_t = R_t - r_t - 1 + R_t / R_{t+1}.$$

The average return from holding long term bonds and borrowing at the t-bill rate is $-.75$ per cent at quarterly rates or -3 per cent at annual rates. Thus bond holders have taken a loss over this sample period in spite of the fact that the average long term rate was 5.9 per cent while the short term rate was only 4.6 per cent. This is a consequence of unexpected increases in interest rates possibly due to unexpected acceleration of inflation.

Maximum likelihood estimation of (18) produced:

$$(24)\qquad y_t = \underset{(-2.2)}{-2.8} + \underset{(1.6)}{.505}\, h_t, \qquad h_t^2 = c_0 + \underset{(2.6)}{.75} \sum_{\tau=1,4} w_\tau \varepsilon_{t-\tau}^2,$$

$$w_\tau = (5-\tau)/10 \qquad (\tau = 1, \ldots, 4),$$

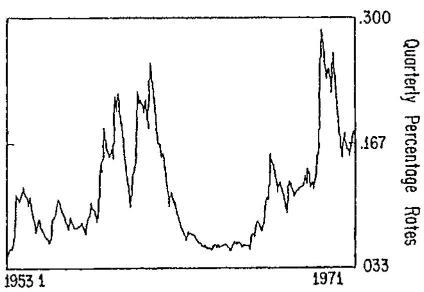

FIGURE 2—Conditional standard errors of one month forward rates.

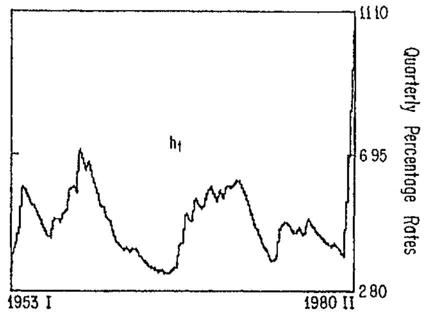

FIGURE 3—Conditional standard errors of quarterly holding yields for Moody's Aaa bond.

for the fourth order ARCH-M model, and

$$\text{(25)} \quad y_t = \underset{(-2.4)}{-3.3} + \underset{(1.9)}{.651}\, h_t, \qquad h_t^2 = c_0 + \underset{(3.4)}{.86} \sum_{\tau=1,12} w_\tau \varepsilon_{t-\tau}^2,$$

$$w_\tau = (13-\tau)/78 \qquad (\tau = 1, \ldots, 12),$$

for the twelfth order model.

These estimates differ from the short end of the spectrum in that they no longer exhibit the explosive ARCH parameter, the coefficient on the risk premium is roughly the same size but has a larger standard error, and the intercept is considerably more negative. When (25) is estimated on data prior to 1978, the coefficient on h_t rises slightly to .84 but the t statistic falls to 1.7. Thus the same model appears to be appropriate; however, the significance falls due to the omission of the highly volatile period of 1979 and 1980. The estimated risk premium is plotted in Figure 3.

Further analysis of these two series is contained in Engle, Lilien, and Robins (1982).

7. CONCLUSIONS

The precision with which agents can predict the future varies significantly over time. In relatively quiet periods, like the mid-1960's, relatively accurate forecasts can be made and agents can speculate on the future without absorbing large risks. In volatile periods, like the early 1970's and early 1980's, forecasts are less certain and speculation is riskier. Risk premia therefore adjust to induce investors to absorb the greater uncertainty associated with holding the risky asset.

In this paper we have extended the simple ARCH technique of measuring conditional variances to the ARCH-M model where the conditional variance is a determinant of the current risk premium, and thus enters the forecasting equation of expected financial returns. Our results from applying this model to three different data sets of bond holding yields are quite promising. ARCH was clearly present in the forecast errors of bond holding yields indicating substantial variation in the degree of uncertainty over time. This measure of uncertainty proved very significant in explaining the expected returns in two of the data sets, and was significant only at slightly more than the 5 per cent level for the third. We therefore conclude that risk premia are not time invariant; rather they vary systematically with agent's perceptions of underlying uncertainty.

While our initial results suggest the promise of the ARCH technique to applications that require the measurement of uncertainty, we feel that the current model is but a first step. The ARCH framework may be applied to more general models of uncertainty and risk. For example, the capital asset pricing model suggests that risk premia are not a function of simple risk, but rather of undiversifiable risk. Risk premia therefore depend on the covariance of asset returns with the returns of the market as a whole. The general ARCH framework may be extended to allow conditional covariances to vary, resulting in time varying risk betas. Such a model is beyond the scope of the current paper and is mentioned to give some indication of possible extensions of our much simpler approach.

Department of Economics, University of California—San Diego, La Jolla, CA 92093, U.S.A.

Department of Economics, University of California—Irvine, Irvine, CA 92717, U.S.A.

and

Trans World Oil, P.O. Box 1252, Hamilton 5, Bermuda

Manuscript received February, 1982; revision received February, 1986.

REFERENCES

AMSLER, CHRISTINE (1984): "Term Structure Variance Bounds and Time Varying Liquidity Premia," *Economics Letters*, 6, 137–144.

BERNDT, E. K., B. H. HALL, R. E. HALL, AND J. A. HAUSMAN (1974): "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 4, 653–665.

BODIE, ZVI, ALEX KANE, AND ROBERT MCDONALD (1984): "Why Haven't Nominal Rates Declined?," *Financial Analysts Journal*, 40, 16–27.

CAMPBELL, JOHN Y. (1984): "Asset Duration and Time Varying Risk Premia," unpublished dissertation, Yale University.

CROWDER, M. J. (1976): "Maximum Likelihood Estimation for Dependent Observations," *Journal of the Royal Statistical Society*, Series B, 38, 45–53.

ENGLE, ROBERT F. (1982a): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.

——— (1982b): "A General Approach to Lagrange Multiplier Model Diagnostics," *Journal of Econometrics*, 20, 83–104.

——— (1984): "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics," *The Handbook of Econometrics, Vol. II*, Chapter 13, ed. by Z. Griliches and M. Intrilligator. Amsterdam: North Holland Publishing Co., pp. 776–826.

ENGLE, ROBERT F., DAVID LILIEN, AND RUSSELL ROBINS (1982): "Estimation of Time Varying Risk Premiums in the Term Structure," UCSD Discussion Paper 82-4.

ENGLE, ROBERT F., DAVID HENDRY, AND JEAN FRANCOIS RICHARD (1983): "Exogeneity," *Econometrica*, 51, 277–304.

FAMA, E. F. (1976): "Forward Rates as Predictors of Future Spot Rates," *Journal of Financial Economics*, 3, 361–367.

MANDLEBROIT, B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394-419.

MANKIW, N. GREGORY, AND LAWRENCE H. SUMMERS (1984): "Do Long-Term Interest Rates Overreact to Short-Term Interest Rates?" *Brookings Papers on Economic Activity*, 223-242.

PANTULA, SASTRY, G. (1984): "Autoregressive Conditionally Heteroscedastic Models," manuscript, Department of Statistics, North Carolina State University.

PESANDO, JAMES E. (1983): "On Expectations, Term Premiums, and the Volatility of Long-Term Interest Rates," *Journal of Monetary Economics*, 12, 467-474.

SARGENT, THOMAS J. (1972): "Rational Expectations and the Term Structure of Interest Rates," *Journal of Money, Credit and Banking*, 4, 74-97.

——— (1979): "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure," *Journal of Monetary Economics*, 5, 133-143.

SHILLER, ROBERT J., JOHN Y. CAMPBELL, AND KERMIT L. SCHOENHOLTZ (1983): "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," *Brookings Papers on Economic Activity*, 173-217.

SHILLER, ROBERT J. (1979): "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," *Journal of Political Economy*, 87, 1190-1219.

——— (1981): "Alternative Tests of Rational Expectations Models: The Case of the Term Structure, *Journal of Econometrics*, 16, 71-87.

WEISS, ANDREW A. (1986): "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107-131.

WOODWARD, SUSAN (1983): "The Liquidity and the Solidity Premia in the Term Structure," *American Economic Review*, 73, 348-361.

SOAH DOCKET NO. 473-21-0538
PUC DOCKET NO. 51415

SOUTHWESTERN ELECTRIC POWER COMPANY'S RESPONSE TO CITIES ADVOCATING REASONABLE DEREGULATION'S THIRD SET OF REQUESTS FOR INFORMATION

Question No. CARD 3-23:

With reference to pages 48-51 of Mr. D'Ascendis' testimony, please: (1) list all regulatory cases (by utility name, docket number, and filing date) in which Ms. Mr. D'Ascendis has provided rate of return testimony and used a non-price regulated proxy group to estimating a market risk premium; (2) indicate all cases (by name, docket number, and date), a regulatory commission has specifically used the equity cost rate results for Mr. D'Ascendis' non-price regulated proxy group approach in arriving at an overall rate of return for a utility; and (3) provide copies of the 'Rate of Return' section of the Commission's decisions for all cases in which a regulatory commission has adopted the equity cost rate results for Mr. D'Ascendis' non-price regulated proxy group.

Response No. CARD 3-23:

1. Mr. D'Ascendis does not use a non-price regulated group to estimate a market risk premium in his analysis on pages 48-51 of his Direct Testimony.
2. Please refer to Mr. D'Ascendis' response to CARD 3-21, part (2).
3. Please refer to Mr. D'Ascendis' response to CARD 3-21, part (3).

Sponsored By: Dylan D'Ascendis Title: Director, ScottMadden, Inc.

Sponsored By: Dylan D'Ascendis Title: Director, ScottMadden, Inc.

SOAH DOCKET NO. 473-21-0538
PUC DOCKET NO. 51415

SOUTHWESTERN ELECTRIC POWER COMPANY'S RESPONSE TO CITIES ADVOCATING REASONABLE DEREGULATION'S THIRD SET OF REQUESTS FOR INFORMATION

Question No. CARD 3-24:

With reference to page 57 of Mr. D'Ascendis' testimony, please provide: (1) the dates and the amounts of equity flotation costs paid by the Company over the 2016-20 time period; and (2) copies of invoices and the associated checks which demonstrate that the Company paid the flotation costs.

Response No. CARD 3-24:

As stated on page 57 of Mr. D'Ascendis' testimony, no analyses regarding flotation costs were performed in this proceeding.

Sponsored By: Dylan D'Ascendis Title: Director, ScottMadden, Inc.

Sponsored By: Dylan D'Ascendis Title: Director, ScottMadden, Inc.

SOAH DOCKET NO. 473-21-0538
PUC DOCKET NO. 51415

SOUTHWESTERN ELECTRIC POWER COMPANY'S RESPONSE TO CITIES ADVOCATING REASONABLE DEREGULATION'S THIRD SET OF REQUESTS FOR INFORMATION

Question No. CARD 3-25:

With reference to pages 56-7 of Mr. D'Ascendis' testimony and Schedule DVD-8, please provide copies of all data, source documents, studies, and analyses used to justify and estimate the small size premium.

Response No. CARD 3-25:

Please refer to the response to CARD 3-15.

Sponsored By: Dylan D'Ascendis Title: Director, ScottMadden, Inc.

Sponsored By: Dylan D'Ascendis Title: Director, ScottMadden, Inc.